CLOUDERA AND TALEND

The Problem:

Enterprises are faced with challenges in handling ever increasing amounts of data. These problems, related to volume and usage demands, have forced IT managers and developers to seek out solutions such as Apache Hadoop for processing to effectively manage this data explosion. Cloudera’s Distribution Including Apache Hadoop (CDH) has quickly become the technology of choice for large organizations in need of sophisticated analysis and transformation of petabytes of structured and complex data. With Talend Integration Suite’s native support of Hadoop, users can access, manipulate and store huge volumes of data in Hadoop and benefit from high-performance, cost-optimized data integration with ultimate scalability.

The Solution:

The Talend Integration Suite (TIS) provides a set of graphical connectors developed to integrate with CDH. Talend provides ETL connectors to extract data from various sources, load data into Hadoop, use the processing power of Hadoop (Pig, Hive, MR), and store the results in HDFS, traditional RDBMS, or other external sources.

By leveraging graphical connectors, customers are able to develop and integrate Hadoop in their existing infrastructure without having an in-depth knowledge of MapReduce or HDFS. Our vision is to serve as the glue between external applications and CDH; TIS achieves this goal by providing 450+ ETL connecting Hadoop with any other well-known application. In addition to providing means of import and export data between HDFS and external sources; Talend provides set connectors that allow implementation of business logic by leveraging the processing power of Hadoop MapReduce.

Talend’s licensing model provides a business advantage to customers that require integration with Hadoop. Talend’s license is not based on number of computers on which it’s deployed or the amount of data that is processed; it is based on number of developers. By structuring our license this way, medium to large enterprises are able to deploy on any number of servers without having to pay additional licensing costs, thus enabling customers to leverage their entire Hadoop cluster.

Key aspects of our Hadoop integration include:

- Native integration with Hadoop by creating direct Java code;
- Basic input/output connectors to HDFS, Hive, HBase, Sequence Files, etc. connecting external sources to Hadoop;
- Largest set of external connectors;
- Hive ELT connectors;
- Graphical transformations of data inside Hadoop with Pig connectors;
- Parallel import/export between Hadoop and RDBMS systems leveraging CDH Sqoop; and
- Affordability through open source model and subscription based licensing.

Talend, Inc.

Industries:
- Data Integration
- Data Quality
- Master Data Management
- Application Integration

Website: http://www.talend.com/

Company Overview
Talend is the recognized market leader in open source data management & application integration. Talend revolutionized the world of data integration when it released the first version of Talend Open Studio in 2006.

Product Overview
Data Integration
- Talend Open Studio, the freely downloadable open source tool.

Data Quality
- Talend Open Profiler, an open source, free data profiling product.
- Talend Data Quality, which includes additional enterprise-grade features.

Master Data Management
- Talend MDM allows organizations to master any data, in any domain without constraint. This unique approach to mastering data unites data integration, data quality, master data and business process management in a single comprehensive solution to meet any master data challenges.
About Talend

Talend is the recognized market leader in open source data management & application integration. Talend revolutionized the world of data integration when it released the first version of Talend Open Studio in 2006.

Talend’s data management solution portfolio now includes operational data integration, ETL, data quality, and master data management. Through the acquisition of Sopera in 2010, Talend became also a key player in application integration. Unlike proprietary, closed solutions, which can only be afforded by the largest and wealthiest organizations, Talend makes middleware solutions available to organizations of all sizes, for all integration needs.